DOCUMENT RESUME

ED 102 539 CS 001 666

AUTHOR Bruning, Roger; Kennedy, Dale
TITLE Some Factors in Children's Identification of Key Words in Written Passages.
PUB DATE Apr 75
NOTE 10p.; Paper presented at the Annual Convention of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Elementary Secondary Education; *Permuted Indexes; Reading; Reading Ability; *Reading Research; *Reading Skills; Semantics; *Vocabulary; *Word Recognition

ABSTRACT

For this study of children's selection of key words, children in the first, third, sixth, ninth and twelfth grades were asked to identify which words, in their judgment, were key words in written materials. Relationship of these choices to skilled readers, to words derived from passage analysis, and to random selections of hypothetical subjects were explored. For matches to skilled readers and to the analytical model, grade level and passage difficulty and length were significant effects. Although the youngest group showed variability of word choice comparable to a purely random selection, further analysis revealed considerable preference for selected grammatical classes in contrast to random choice. Recall protocols after one week and six months showed vocabulary usage frequency parallel to initial key word identification frequency. (Author)

ED102539

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Some factors in children's identification of key words in written passages

Roger Bruning and

Dale Kennedy

University of Nebraska, Lincoln

One initial step toward the comprehension of a written message is the identification of those vocabulary entries which are critical to the understanding of the passage. The selection of such words would seem to hinge on a variety of interacting skills acquired over the course of increasing experience with both the semantic and syntactic features of the language. The present study involved an exploration of this identification process in children of various ages. In general, it was hypothesized that the younger children would more nearly approximate a random selection model, particularly in more difficult or longer passages, while older children would tend to converge in their choices toward those of skilled readers, performing within the limits of their linguistic sophistication.

Examination of the children's selection of key words was made in relationship to three reference models. The first, the Skilled Reader model, was empirically derived and operationalized as the highest ranking key word choices of a graduate student group of presumed skilled readers. At the opposite end of the continuum, a Random model was similarly empirically-based and was represented by randomly determined word selections made by hypothetical subjects from the reading passages. A third model was based on characteristics of the passage itself and derived from passage analysis. Called a Multiple Reference model, it focused on repeated lexical items appearing in the passages. Attention to referential connections is one feature of a system for structural analysis of prose recently proposed by Calfee, et. al. (1974). In the Calfee reference system, the most frequently occurring words, both in direct and indirect reference, form the nuclei of topics. In the present study, the student's choices as key words were compared to form class items which occurred most frequently in the passages.

Of interest to the present investigators were the relationships of key word identification skills to reading comprehension skills and to recall. To what extent does the child's ability to identify key words in written passages relate to more standard measures of language sophistication? For purposes of the present study, scores on the key word identification task (matches to the Skilled Reader and Multiple Reference models) were correlated with scores on subtests of standardized achievement tests in the areas of reading comprehension and vocabulary. The final component of the present study was the testing of recall. To what extent are there relationships between key word identifications and the vocabulary appearing in the recall protocols of children after a period of time? The general expectation here was that the lexical entries identified as

CS 001 666

key words would be likewise present in any recall attempts related to the passages, to the extent that the key words described the central topical units for each of the passages.

Method

Subjects

The research participants were 135 1st, 3rd, 6th, 9th, and 12th grade students from a small rural public school,[1] together with a reference group of 12 graduate students in a department of Educational Psychology. To facilitate comparison across grade level, the data from one to four students at each grade level were randomly discarded, leaving the data for 24 students at each of the five grade levels for analysis.

Procedure

All research participants were presented with identical reading materials and instructions. The materials consisted of four paragraphs varying on the dimensions of difficulty and length. The passages were 25, 28, 49 and 40 words in length and are presented below:

Easy: The kitten was sitting on the step. It was warm and the kitten began to feel sleepy. It closed its eyes. Soon it was asleep.

Difficult: In the early frontier days a girl's education consisted of learning how to sew and embroider clothing, and how to prepare food for themselves and the working men.

Easy: Elephants eat many kinds of plants, but they never eat other animals. They are so big that they eat most of the time. Elephants like to eat the small leaves that grow at the tops of the trees. Sometimes these leaves are too high for even an elephant to get.

Difficult: The various codes which were presented to you at Crossgates . . . contradicted one another if you worked out their implications. The essential conflict was between the tradition of 19th-century asceticism and the actually existing luxury and snobbery of the pre-1914 age.[2]

Each paragraph was typed, double-spaced, on a single page of a four-page booklet, with pages randomly ordered for each subject. The students worked in grade-level groups beginning with half of the first graders, followed by the 12th graders, 6th graders, the second half of the first grade, 3rd graders, and 9th graders. The students were told to read each paragraph and to pick out the five words which, in their judgment, were the most "important words" in the paragraph. Half of the students at each grade level were instructed to indicate their responses by underlining (Underline group), while the other half wrote the words in blanks which were placed immediately under the paragraph (Write group). The latter task, it was hypothesized, may place more rigorous demands on

memory, resulting in increased variability, particularly in younger readers. To avoid possible confusion, the first graders were instructed and run in separate groups for Underline and Write conditions. To all students, the experimenters emphasized that the task was not a test, and there were no right or wrong choices, and that students should only choose those words they personally considered most important. The graduate student reference group was given identical instructions to those given to the elementary and secondary pupils and responded to materials under the Write condition.

The data on word frequency and choice variability were provided by means of a computer program developed by Kennedy (1974) for analysis of verbal protocols. The extent to which subjects matched the Skilled Reader model was operationalized as the number of words selected within each paragraph which matched the five most frequent choices of the graduate student group. These scores were analyzed in a 5 x 2 x 2 x 2 analysis of variance, with factors of grade level and response mode as between-subjects variables and paragraph length and difficulty as within-subjects variables.

The passages were analyzed for instances of multiple references, either direct or indirect, of substantive, nonfunction terms. In the short/easy passage, kitten/it/its has six entries and sleep/asleep two. In the short/difficult passage, education/learning has two entries. Elephants/elephant/they appears four times, eat four times, and plants/leaves three times in the long/easy passage while, in the long/difficult passage, codes/one(another) appeared three times and contradicted/conflict twice. These selections formed the basis for the Multiple Reference model analysis.

Matches to this latter model were possible for the written response groups only, in which word choice order variability was possible. To derive a score for each individual, the initial entry of each multiple reference in the passage was noted. Matches to the initial reference item for each multiple entry set were made, using a like number of subject-selected words to determine the extent of match to the model. (e.g., whether codes and contradicted appeared as the first two words of a subject's selections for that passage)

Third, sixth, 9th and 12th grade children's matching scores to each of the two models were subjected to correlational analysis, comparing them to the Vocabulary and Reading Comprehension subtest scores of the Iowa Test of Basic Skills (Grades 3 and 6) and the Iowa Tests of Educational Development (Grades 9 and 12).

An unannounced recall followup was subsequently made for 3rd, 6th, 9th, and 12th grade pupils six months following the initial encounter with the passage. All earlier participants in the key word identification segment were presented with four sheets of lined paper, reminded that there were four passages that they had read, and were asked to write as much as they recalled of each of the four passages. In addition, a new group of sixth graders was presented with the materials and asked only to read them carefully. A short "feedback" session of approximately two to three minutes duration soliciting the children's reactions to the

passages was given as the ostensible purpose of this session. One week later the classroom teacher asked these children to write what they recalled in the same fashion as the first groups had previously been asked. The recall requirement in this instance, as in the earlier attempt, was unannounced.

Results

The total number of different words selected by grade level and condition pooled across response mode are presented in Table 1, together with the mean matching scores of subjects relative to the Skilled Reader model

---

Insert Table 1 about here

Significant effects ($p < .01$) were present for grade level ($F = 55.89$, $df = 1/110$), length of the passage ($F = 31.59$; $df = 1/110$) and difficulty ($F = 340.88$, $df = 1/110$). The interactions of grade level and difficulty ($F = 14.09$, $df = 4/110$) and response mode and difficulty ($F = 5.78$; $df = 1/110$) were also significant. No other effects were significant.

With regard to the Multiple Reference model, only scores of the Write condition were analyzed, since response ordering was meaningful only in this condition. When matching scores were analyzed, significant ($p < .01$) differences were present for the main effects of grade level ($F = 9.79$; $df = 4/55$), for passage length ($F = 120.58$; $df = 1/55$) and for difficulty of the passage ($F = 187.92$, $df = 1/55$). No interactions were significant. As in matches to the skilled readers, the matching scores improved with increasing grade level, and deteriorated with increasing length and difficulty of the passage.

In general, the data reveal a developmental trend continuing through all of the school years of convergence in the identification of important words in prose passages and show this selection to be quite strongly affected by the variables of difficulty and length of passage. Students' selections conformed increasingly across increasing grade level to those of skilled readers and increasingly matched the product of an analysis based upon conceptual redundancy in the materials.

However, even the youngest children, although they showed nearly as large variability as a group in their selections as a hypothetical random selection group (See Table 1), tended to strongly reject function words and pronominalizations in favor of noun, verb, adjective and adverbial classes. Hypothetical Random Model subjects selecting words on a purely random basis from the passages showed an overall proportion of selection of words in the latter categories of 55%, significantly lower than the 88% selected by the first grade children ($p < .01$). Interestingly, almost all selections of

function words by the 1st graders (13 of the 15 words chosen in this category) came from the Write group, where sequential processing may have been interrupted, perhaps with resultant loss of intrasentential cues available to the Underline group. Older groups, probably less context bound, and certainly more sophisticated linguistically, were generally unaffected by the response mode variable, choosing almost exclusively in the form class categories. Percentages of words selected in substantive categories for the 3rd, 6th, 9th and 12th grades levels were 99, 100, 100, and 99 percent respectively.

Standardized achievement test scores were available for grades 3, 6, 9, and 12, and correlational analyses were performed between matching scores and two subtests. Generally low correlations were obtained between matching scores and the two selected achievement test variables, reading comprehension and vocabulary knowledge. No significant correlations between either matches to skilled readers or matches to the analytical model and the standardized test scores were evident at grades 6, 9, or 12. However, significant relationships to matching scores for the Multiple Reference model were present for both the third graders' vocabulary ($r = 0.71$, $p < .01$) and for third graders' reading comprehension scores ($r = .51$, $p < .05$).

Recall of the passages was assessed in a read only sixth graders' group after an interval of one week and in the key word identification group of sixth graders after approximately six months had passed. After an interval of six months, recall among the total group of subjects consisted of only a few scattered words (See Table 2); after a week's time, however, forgetting

---

Insert Table 2 about here

---

was much less complete, but nonetheless substantial, particularly in relationship to recall of any detailed information in the passages. Ranks of the original key word identification of the sixth grade group for each of the passages are presented in Table 2, together with the ranks of occurrances of these words in participants' recall protocols after one week and 6 months delay.

Discussion

A particularly interesting feature of the present data is the close correspondence of the initial rankings of key words in the passages with the rankings of frequency of occurrance of these same words in recall protocols after one week and after a six month interval. In the present materials, at least, the key features retrieved from memory are the same vocabulary items identified in the key word identification process as those terms most "important" in the passage. In the case of the six month recall attempt there was, of course, a potential for confounding of the key word

identification and recall, since the recall was assessed in the same group which had earlier chose those words they considered most important in the passage. Subsequent memory, therefore, may have been a function of the identification task to some extent. The potential confounding, however, was not a factor in the one week recall group of sixth graders, who read for comprehension only, with no instructions to focus on key ideas.

One initial consideration in the present study was to look at the feasability of an assessment process which employed the key word identification procedure. Such a procedure would have the advantages of permitting a large number of responses to a variety of passages in a short period of time, yielding an easily scorable set of responses, and potentially, providing insight into such variables as children's syntactical knowledge, logical processes, and idiosyncracy of responding. However, while responses showed convergence with presumed increasing sophistication with the language, matching scores were only weakly related for the most part to standardized test performance in vocabulary knowledge and reading comprehension. The potentially interesting relationship between key word identification patterns and passage vocabulary knowledge and passage comprehension was not explored in the present study and should be investigated in further research.

The results on the memory portion of the study generally support hypotheses emphasizing the role of comprehension in processing on subsequent recall (e.g., Barclay, 1973; Jenkins, 1974). Where there was no comprehension, there was little or no recall. Although the children studied the Crossgates passage as intently as the others, there was virtually no remembering of it at either interval by any group, including the high school seniors. At the same time, the nearly complete lack of recall of any passage's content after six months was somewhat surprising. However, the nature of the key word identification process may be that it does not necessarily require "comprehension" of the passages in the most complete sense of the term. That is, identification of key vocabulary may be accomplished with only partial attention to meanings present in the passage and may, in fact, divert attention away from associative and other cognitive processes ordinarily present in responding to meaningful written materials.

References

Barclay, J. R. The role of comprehension in remembering sentences. Cognitive Psychology, 1973, 4, 229-254.

Calfee, Robert, Filby, N. and Drum, P. A system for structural analysis of prose. A paper presented at the American Educational Research Association Convention, Chicago, April, 1974.

Jenkins, J. J. Remember that old theory of memory? Well, forget it! American Psychologist, 1974, 29, 785-789.

Kennedy, Dale An approach to the theory and measurement of language usage patterns. A paper presented to the Nebraska Academy of Sciences, Lincoln, Nebraska, April, 1974.

Footnotes

[1] The authors wish to thank Mr. LeRoy Garrels and Mr. George McCabe of Friend, Nebraska Public Schools for their assistance in making arrangements for the study.

[2] The first-three passages were written by the experimenters and are related to passages appearing in published tests of comprehension. The last passage is from an essay by George Orwell, "Such, Such were the Joys...." in A Collection of Essays by George Orwell, (Garden City, New York: Doubleday and Company, Inc., 1954)

Table 1

Number of Different Words per Group (N = 24) and mean Matching Scores of the Subject's Five Words to Selections of Skilled Readers

	Passage Characteristics							
	Short				Long			
	Easy		Difficult		Easy		Difficult	
Condition	Number of Different Words	Mean Matching Score	Number of Different Words	Mean Matching Score	Number of Different Words	Mean Matching Score	Number of Different Words	Mean Matching Score
Hypothetical Random	18	1.33	23	0.67	33	1.00	31	0.58
Grade 1	17	2.12	22	1.54	33	1.58	27	0.92
Grade 3	10	3.50	16	1.83	18	3.46	18	1.21
Grade 6	9	3.63	13	2.96	14	3.67	18	1.96
Grade 9	10	3.63	15	3.21	14	3.54	19	2.20
Grade 12	10	3.96	16	3.04	11	3.71	17	2.71

Table 2

Rank of key word choices and appearance of these words in one-week and six-month recall attempts of sixth graders

	Initial Key Word Choices[1]		One-Week Recall[2]				Six-Month Recall[1]			
			Verbatim		Verbatim + Transforms[3]		Verbatim		Verbatim + Transforms[3]	
Passage	Rank	Word N	Word	N	Word	N	Word	N	Word	N
kitten	1	SLEEPY 22	kitten	22	*kitten*	24	kitten	4	*kitten*	8
	2	KITTEN 20	asleep	19	*asleep*	19	step	1	*step*	1
	3	WARM 18	warm	9	*warm*	18				
	4	ASLEEP 16	step	7	*step*	11				
	5	STEP 11								
frontier	1	FRONTIER 19	girl	13	*girl*	18	girl	1	*girl*	2
	2	EDUCATION 18	frontier	9	*frontier*	18				
	3	EMBROIDER 15	sew	7	*sew*	9				
	4	GIRLS 12	education	4	*education*	9				
	5	SEW 11	embroider	3	*embroider*	3				
elephants	1	ELEPHANTS 21	elephants	21	*elephants*	23	elephants	3	*elephants*	3
	2	LEAVES 20	eat	21	*eat*	21	animals	2	*animals*	2
	3	PLANTS 17	plants	12	*plants*	12				
	4	ANIMALS 17	leaves	10	*leaves*	10				
	5	EAT 14	animals	5	*animals*	5				
crossgates	1	CROSSGATES 18	crossgates	4						
	2	CODES 14	codes	4						
	3	CONTRADICTED 13					No references were made in six-month recall to the Crossgates passage.			
	4	IMPLICATIONS 11								
	5	ASCETISM 11								

1. This first group of sixth graders (N = 24) identified key words; recall was then tested after six months.
2. The second group of sixth graders (N = 24) was tested for recall one week after reading for comprehension.
3. Transforms were kitten (cat, squirrel), warm (hot, heatness), step (stairs, porch), frontier (pioneer, colonial, olden days), girl (women, ladies), sew (knit, mend), education (learn, teach) and elephants (dinasour, bear).